DOCUMENT RESUME

ED 306 243 TM 013 010

AUTHOR Nagy, Philip; Moorhead, Roslyn
TITLE Administrative Response to Classroom Testing Data: A Problem Solving Perspective.
PUB DATE Mar 89
NOTE 32p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, March 28-30, 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Administrator Attitudes; Case Studies; Classroom Environment; *Classroom Techniques; Elementary Secondary Education; *Evaluation Methods; Grading; *Principals; Problem Solving; School Administration; *Student Evaluation; Testing Problems; *Test Results; Test Use; Verbal Ability

ABSTRACT

To study the responses of school administrators to classroom testing and problems associated with it, practicing principals were presented with a case study. The case contained information on a district-wide report outlining inconsistencies in teachers' evaluation practices and a parental complaint about the evaluation practices of a staff member. The 15 principals were selected for convenience, and represented a slightly more skilled sample than a true random sample. Response protocols were segmented into thought units for analysis in one of eight areas. With one low outlier and two high outliers, responses clustered closely around segments in the following areas: (1) student; (2) teacher; (3) parental complaint; (4) study; (5) department; (6) school; (7) district; and (8) methods. Statements were categorized according to type--reasoning, action, reiteration, values, background, summary, and comment. There was little evidence that the statement types and patterns of explanation bore a relationship to the quality of action individuals might take. The study may have been tapping differences in verbalization skills rather than school administration skills. Reasonably consistent desires were found among the principals for grading consistency and some degree of consistency in administrative ability. Finally, only about half of the subjects actually related the case study to the parental complaint. A table and two figures illustrate the data. Appendix A presents the case study and Appendix B gives the principals' value statements. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED306243

Administrative Response to Classroom Testing Data: A Problem Solving Perspective

Philip Nagy
The Ontario Institute for Studies in Education

and

Roslyn Moorhead
The University of Western Ontario

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

PHILIP NAGY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, 1989.

TM013010

Introduction

The intent of this report is to set the issue of classroom testing and problems associated with it in the broader context of the school program and its ongoing maintenance. To accomplish this end, we presented practising high school principals with a case study, and examined their responses to the situation presented. The case contained information on a district-wide report outlining inconsistencies in teachers' evaluation practices, as well as an immediate parent complaint about the evaluation practices of a staff member.

In addition to dealing with the substance of the principals' reactions, we also provide some detail on our methods, from cognitive science, of examining the transcripts provided by our sample. Apart from questions involving this specific data set, we hope that our methodological comments are relevant to a broader issue in measurement, that of the systematic analysis of complex student achievement data. In particular, and very much in keeping with a recent article by Perkins and Salomon (1989), we began our analysis from a general skills perspective, and found ourselves needing to add context-specific elements to complete our task. The reasons for these modifications shed light on the analysis of problem solving data, whether the subjects be students, teachers, or principals.

Background

The rationale for this line of enquiry rests on three points: first, inconsistencies in classroom testing practice are a serious problem; second, the principal is a key figure in contributing to the solution of this problem; and third, our analysis methods, from the literature on problem solving, are appropriate for the task we have set ourselves.

Testing Practice. Measurement specialists have begun to make teachers' assessment practices a target for examination. In an example of this line of

enquiry, Traub, Nagy, MacRury and Klaiman (1988) have examined the assessment practices of high school calculus teachers, all from large, relatively wealthy school districts in the urban core of Southern Ontario. The teachers kept daily logs of their teaching and evaluation activities, and provided copies of their tests, quizzes, and exams for an entire semester. In addition, they all marked the same set of student exams obtained from a school outside the study and provided by the research team. The results of the investigation showed considerable variation in teacher practice. Teachers differed substantially in use of classroom time, in the amount of student effort required, in relative emphasis given topics, in match between content taught and content tested, in methods used for assessing student achievement, and in overall generosity in awarding marks.

As well, Traub et al have provided a convenient, although brief review of related work. They report:

-- Teacher-developed tests and classroom observations are the most common methods of assessing students (Herman & Dorr-Bremme, 1983);

-- Variations in assessment practices exist across subject areas and grades (Green & Stager, 1986; Terwilliger, 1987): For example: elementary teachers tend to favour observations and secondary teachers, testing (Stiggins & Bridgeford, 1985); and mathematics is heavily test-oriented compared to other subject areas (Wahlstrom & Danley, 1976).

-- Teachers need training in teaching and assessing higher-order thinking skills (Haertel, 1986; Stiggins, 1988); and

-- The measurement community needs to learn more about how assessment is conducted within the classroom by using more focussed research methods, such as classroom observation (Stiggins & Bridgeford, 1985; Stiggins, Conklin & Bridgeford, 1986; Anderson, 1987).

The Traub et al results were incorporated in a case study presented to a sample of principals in this study. The case study (Appendix A) contains a summary of the results obtained by Traub et al, modified only by changing the number of schools involved from 17 to 15, and by placing these schools in one school district rather than 13.

The Role of the Principal. Current administrative theory assigns the task of curriculum leadership and management, including student evaluation practices, to the principal. Rutherford (1985) has outlined concrete aspects of effective school leaders. Among his indicators of effectiveness are the setting of clear expectations for both teachers and students, knowledge of individual classrooms, instructional support, and monitoring of progress. Leithwood and Montgomery (1986) identify four levels of principal effectiveness, which focus, from lowest to highest, on the rules, the climate, the program, and the students. Among the factors under the principal's rather than the classroom teacher's control, they include assessment, recording, and reporting procedures. At the policy level, the expectation of the Ontario Ministry of Education (1988), stated in their recent manual on curriculum management, is that the principal will assume responsibility for achieving the goals of the curriculum. From both the theoretical and policy perspectives, a major role of the principal involves monitoring of the degree to which the objectives of the curriculum have been accomplished, i.e., student evaluation, and assisting teachers in the improvement of their practice.

Problem Solving. Our strategy for analyzing the data collected in this study is rooted in the artificial intelligence approach to how people solve problems. Problems can be placed on a continuum from well-structured (logic puzzles, chess, and physics problems) to ill-structured (economic, social, and political problems). Well-structured problems are characterized by low linguistic load, little reliance on context, and widely accepted procedures and

correct answers. Ill-structured problems are characterized by high linguistic load, much reliance on context, and debate about accepted procedures and solutions. Our focus, principals' responses to case studies, is an example of an "ill-structured" problem.

The most immediate source for our analytic methods is the work of James Voss and colleagues at The Learning Research and Development Center, University of Pittsburgh (Penner and Voss, 1983; Voss, Green, Post and Penner, 1983). They analyzed responses to protocols on the question of (ill-structured) Soviet agricultural production problems, using parallel structures for problem solving and verbal reasoning. Examples of steps ("operators") in the problem solving structure include "state subproblem," "provide support," and "summarize." In the verbal reasoning structure, they categorized subjects' responses as, for example, "state assertion," "elaborate," and "state qualifier."

Earlier work (Nagy and Allison, 1988), involving responses from a small sample of experienced principals to a different case study, provided an initial adaptation and refinement of the Voss et al methods. In that analysis, Nagy and Allison first categorized statements from protocols into a set of 19 categories (e.g., reasoning, subproblem, summary). They then drew diagrams in which statements were grouped by subproblem; a system of indentations showed subordinate relations between statements. For example, a reasoning statement was indented once from the action it justified, and a subsequent elaboration of that reasoning was indented a second time. Finally, they drew highly simplified summary diagrams showing the pattern of subproblems identified and actions taken.

In the present analysis, we began with essentially the same system. We found, however, that the detailed statement categorization caused us to dwell more on the language of the respondent than on the substance of his/her response to the case and that our category system failed to provide a basis for

judgments of quality (e.g., for differentiating poor problem solving from better). Thus, we greatly simplified the statement categorization system, and paid more attention to the subproblems addressed by the principals and the specific actions that would have been taken. In retrospect, we question the extent to which meethods involving categorization of statement types tap verbalization ability compared to problem solving ability. On this issue, we have some suggestions for refinement of the methodology.

Method

Sample. The principals who participated in this study were chosen for convenience -- essentially from personal knowledge and for geographical location. As about half came to our attention through supervisor or peer recommendation, the sample should probably be considered slightly more skilled than a random sample of experienced, practising principals. However, apart from knowing that three of the 15 are in their first or second year at the principalship, we have no systematic way of distinguishing relative expertise. All interviews were conducted in the fall of 1988, in the offices of the principals, and at their convenience.

The Case Study. The case study (Appendix A) contains two elements. The first element involves a results of a research study recently completed in the high schools of the district. The study, using one calculus teacher from each school, showed large differences (a) in use of class time (e.g., seatwork versus direct teaching), (b) in amount of practice given, (c) in content covered, (d) in amount of evidence collected for grade assignment, (e) in match between content taught and content tested, and (e) in grades assigned to a common set of student exams. The study was reported to the board, where it occasioned some comment. The principal in the situation had previously asked his math head and department to discuss it, and they reported no problems in

their school. The second element of the case study is a parental complaint. The student in question, reported by her father as having done well in math in the past, is depicted as having difficulties in calculus. The father wants a change from one teacher to another. In the scenario, the responding principal is asked to place him/herself in the shoes of the principal in the case study, and describe what he/she would do in the situation. The interviewer answered clarifying questions, but no probes were used to explore areas that did not spontaneously occur to the responding principal.

For context, calculus is taught in Ontario high schools only in the final, the OAC (Ontario Academic Credit) year. For most students, this is the fifth year of high school. However, the system is designed so that the more able, if they are willing to take a "no-frills" curriculum, can finish high school, with university entrance requirements complete, in four years. For further context, there are no provincially-set exams nor any equivalent of the SAT in Ontario. High school grades are the major basis for university entrance, and they are determined entirely by the teacher.

The Analysis. In the initial analysis, response protocols were segmented into thought units and numbered sequentially. A "thought unit", intuitively defined, was essentially a sentence; compound sentences were split, and items listed in a sentence were individually numbered. Next, these units were categorized into about 20 categories using a system adapted from Nagy and Allison (1988). Finally, diagrams were drawn to show the flow of the problem solution. In these diagrams, flowchart-like symbols for statement types were used, sequence was shown by vertical (downward) shifts, and change in sub-problem area by left-to-right shifts.

As described above, this approach was judged unsatisfactory because it dealt too much with the articulation of the problem solving, and not enough with the substance. We felt that we were dissecting differences in oral

communication ability rather than problem solving ability. More importantly, it did not provide a systematic method of describing what each individual chose to do or not do in the presented situation. It left the distinction between conscious linking of subproblems and uncritical meandering between subproblems to the unaided subjective judgement of the analyst.

The revised analysis proceeded from two perspectives: that of the problem solving processes used, and that of the issues addressed. Two major changes from the first attempt were undertaken. First, the preliminary analysis had given us a sense of the range of areas addressed and actions taken. We made a decision to assign each statement uniquely to one of eight areas. (This required some re-segmentation into much smaller thought units.) Second, the number of statement types was greatly reduced, from about twenty to just eight. Mainly, a large variety of subtle distinctions were ignored in favour of one larger category, "reasoning."

The analysis reported has four components: first, the emphasis given to each area; second, what was actually accomplished by each respondent; third, the statement types used; and fourth, the explicit value statements made by the respondents. Where possible, we attempt to relate these different views of the data to each other.

Results

The responses were segmented into statements. With one low outlier (Principal #1, 23 segments) and two high outliers (#4, 461 segments, and #6, 259 segments), the responses clustered quite closely around 100 segments. Preliminary analysis showed that the respondents dealt with eight different areas:

STUDENT -- gathering information on the student, talking with her, or setting up assistance.

TEACHER -- talking to the teacher, or getting information on the teacher's performance (also includes any dealing with the receiving teacher).

PARENT/COMPLAINT -- talking to the parent, trying to resolve the issue; excludes any information gathering needed for resolution of the conflict.

STUDY -- discussion of the study of assessment practices in the abstract; that is, not in the context of what the department, the school, or the district ought to do.

DEPARTMENT -- talking to the four teachers of calculus as a group, rather than the teacher whose student has complained; or talking to the department head; discussion of procedures the department ought to have in place, or information that ought to be available.

SCHOOL -- discussion of school-wide issues related to the case; response had to involve explicitly other departments beyond the math department, or clearly talk of "departments" or "heads" in the plural.

DISTRICT -- discussion of issues involving the elected board, the district central administration, or district level committees.

METHODS -- any discussion of how to solve problems, or how to deal with conflict.

Figure 1 shows the degree of emphasis in each principal's protocol on the eight areas. The sum of the widths of the boxes in Figure 1 for each principal is a constant, representing 100% of the statements given in response to the case. As can be seen, while all respondents paid some attention to the parent making the complaint, only 2/3 of them talked directly to the student or the teacher. Only 1/3 dealt with issues beyond the school boundaries, and of these (not shown in the figure), only one made a reference beyond the school district (i.e., to the effect that students had to compete with students from other schools for admission to university). Just over one-half of the respondents

discussed issues raised in the research study apart from the specific parental complaint, while only two dealt in a meta-analytic fashion with problems or conflict. All but one of the respondents discussed departmental procedures. A few of the respondents (#3, 14, 15) can be seen to have addressed only a limited number of the eight areas, but most dealt with five of the eight.

Table 1 gives a matrix of what was accomplished or addressed by each principal. These accomplishments do not have a one-to-one correspondence with the actions taken; several actions might be undertaken for one accomplishment, for example, to find information on the student. As well, there are only seven categories in Table 1, compared to eight in Figure 1; no aspect of the discussion of the study could be considered an accomplishment apart from accomplishments involving the teacher, the department, or the school. An examination of Table 1 shows that only six of the 15 responding principals would have checked to see if the information from the parent on the student was accurate. However, eight of 15 set up some type of assistance, either academic or personal, for her. Except for two respondents, there was generally less investigation of the individual teacher than of the student. Only about one-half the sample tried to find out what the department did to monitor or control teacher practices with respect to the topics examined in the district study. In effect, such investigation shows a clear linkage between the individual complaint and the general issue. Thus, one-half the sample treated the study and the complaint as only marginally related.

The information in Table 1 sheds light on how the principals felt about the issues raised in the reported study of calculus teacher practices. Three respondents talked of dealing with the specific complaint within the school. Two of these individuals were concerned to avoid the imposition of a solution (to the general issue rather than the complaint) from outside the school, while the third talked of the political ramifications of having a hostile parent go

to the district administration. Only three respondents stated that the district office needed to be involved in resolving between-teacher inconsistencies.

To further explore how respondents felt about the issues, we isolated their value statements (see Appendix B). About half stated that the promotion of uniformity of marks within the school was important, but only two spoke specifically in favour of common exams, even within the school. Most talked of ongoing communication among teachers as the main vehicle for the promotion of uniform marking standards. Also, in only a couple of cases is it explicit that the motivation for achieving consistency is fairness for students rather than the avoidance of complaints. (More on this below, where we argue for development of a systematic set of probes for this type of work.)

As a final step in our analysis, we categorized statements according to type:

REASONING -- a general term including several categories of the original analysis: identifying subproblems, pointing out assumptions, citing information from the case, identifying missing information, predicting, explaining an action, analyzing the case; DETAILS of Reasoning statements are tabulated separately.

ACTION -- specific statements of what the respondent would do; again, DETAILS of Action statements are tabulated separately.

REITERATION -- simple repetition, which was distinguished from attempts to provide a summary.

VALUES -- statements of what the respondent considers important, about assessment, students, or education; where detail or elaboration was provided, it was not tabulated separately.

BACKGROUND -- relating the case to the respondent's personal experience; in two situations, a respondent either began or ended the discussion with a lengthy personal anecdote; these were simply not analysed.

SUMMARY -- summaries of previous steps.

COMMENT -- statements about the process itself, including both throw-aways ("Where was I?") and substantial meta-analytic statements ("I like to lay out the steps in a problem solving process before I begin").

Figure 2 gives a breakdown of the statement types used by each respondent. Again, the sum of the width of boxes for each respondent is a constant, representing 100% of the statements. For Reasoning and Action, the bar down each box distinguishes Reasoning or Action, on the left, from Detail, on the right. In the two cases with no bar, there was no detail provided. One interesting comparison to make in Figure 2 is between Reasoning and Action. For about one-half the sample, there were more than twice as many Reasoning statements as Action statements.

There are several other patterns within Figure 2, but their relationship to the quality or nature of response to the case is not clear. Voss and his colleagues suggest in their research that at least two factors differentiate the expert from the novice: experts isolate more abstract or general problems as being the primary factors producing a given problem, and experts show more skill in providing supportive reasoning for their stated solutions, in stating subproblems which may emerge from the solutions and in evaluating solutions in relation to constraints. Neither of these patterns is apparent here, however.

Subproblem identification in our protocols was frequently implicit rather than explicit. None of the respondents consistently articulated or defined subproblems associated with the case; rather, subproblems were implicitly identified in the stated solutions or action plans. Hence a category for subproblem identification has not been included as one of the seven statement types. Instead our analysis focussed on the matrix of activities and the areas commented on by the respondents. We found that everyone dealt with the immediate problem of responding to the parent, but only half dealt with the

broader or more abstract issues concerning the study or the ramifications for the school, and only one-third mentioned the school district. However, of the three novice principals (those with less than three years experience), one (#2) talked of all three areas, and another (#4) talked of two of them.

In terms of reasoned arguments, some respondents reiterated and backtracked more than others; some provided periodic summaries; others related their actions strongly to their own school situation; and some expressed many more values than others. Again, however, there was no correlation between the extent of these activities and the experience of the principal. In fact, a novice (#4) was one of those who employed substantial supportive reasoning. Moreover, in our subjective opinion, more reasoning is not necessarily better; some respondents engaged in critical analysis of the problem and others in uncritical meandering. The present analysis by statement types does not differentiate between these two.

Discussion

Several issues, both methodological and substantive, arise from this investigation. With a reminder that we believe the analysis of complex problem solving data should be of interest to the measurement community, we will begin with comments on our own methodology.

First, there is little evidence in the present data set to suggest that statement types used and patterns of explanation bear a relationship to the quality of action the individuals might take when placed in a situation similar to the presented case. We are at this point unable to relate differences in ability to articulate one's views to ability to run a school successfully. Similarly, there is little reason to believe that patterns of statement use, such as in Figure 2, can be related to substantive issues involving student assessment attitudes and practice. One explanation, which we are currently

pursuing, is that we have a restricted-range problem. Research on expert-novice differences (e.g., Patel and Groen, 1986) has shown differences in expert use of "forward reasoning" and novice use of "backward reasoning". By such a criterion, all our sample were experts. We may indeed find the data on statement type useful in differentiating between responses from practising principals and true novices such as student teachers, or less experienced classroom teachers.

Second, we felt the need to introduce Table 1, which amounts to a subject specific "marking scheme", the subject being school administration. This is in keeping with Glaser's (1984) argument for the importance of specific knowledge over general skills. Using a content-based system, we can distinguish levels of success at the task. However, as Perkins and Salomon (1989) point out, more atypical problems might better allow differentiation between experts and novices in general skill (i.e., the approach in Figure 2) rather than subject-specific skill (i.e., the approach in Table 1). Perhaps our case study was not novel enough for our sample. At the time, we found comments such as "this is a real problem -- it happened last week" quite gratifying, showing that we had not lost touch with school realities. However, a truly novel problem may have allowed differentiation in terms of general skills as well as subject-specific accomplishments.

Third, we felt discomfort that we were tapping differences in verbalization skills rather than school administration skills. This suggests development of a substantial set of probes to compensate for verbalization ability. With such probes, we would still be able to distinguish what is spontaneously offered by respondents, but we would be less in the dark about all the things not said. That is, we would know if a given respondent had actively chosen not to take some action, or had merely forgotten to mention it in the discussion. As Nagy and Allison (1988) pointed out, our experts are

experts at doing things, rather than, as were Voss' experts, at talking about things. However, this argument cannot be carried too far; one might reasonably argue that principals earn their living with words as much as do academic agricultural experts.

On the substantive side, we found reasonably consistent desires among our sample for grading consistency, but large differences in how large a role to play in achieving this consistency. Only a minority took an active hand in departmental discussions; most left it to the department chairman. This view is not in keeping with the view of the principal as curriculum leader. However, it does seem to represent current practice in Ontario high schools.

Second, and this may be a methodological point poorly disguised, we have tapped the "craft" of administration. Our subjective impression is that our group of respondents, all successful by different criteria, actually differ less in administrative ability than our results show. This suggests that we need to explore the importance of rationality -- the requirement for success at our task -- to actually running a school.

Finally, and perhaps most important for classroom assessment, only about one-half our sample actually treated the study described in the case study as relevant to the parent complaint. This is perhaps the most important food for thought for those involved in the improvement of teachers' assessment skills.

References

Anderson, J. O. (1987). Teacher practices in and attitudes towards student assessment. Paper presented at the Annual Meeting of the Canadian Educational Researchers' Association, McMaster University, Hamilton, Ontario. June, 1987.

Glaser, R. (1984). Education and thinking: The role of knowledge. American Psychologist, 39, 93-104.

Green, K. E., & Stager, S. F. (1986). Effects of training, grade level, and subject taught on the types of tests and test items used by teachers. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Herman, J. L., & Dorr-Bremme, D. W. (1983). Uses of testing in the schools: A national profile. New Directions for Testing and Measurement, 19, 7-17.

Leithwood, K.A., & Montgomery, D.J. (1986). Improving principal effectiveness: The principal profile. Toronto: OISE Press.

Nagy, P., & Allison, P. (1988). School-level problem solving: A cognitive perspective. A paper presented at the Annual Meeting of the Canadian Society for the Study of Education, Windsor.

Ontario Ministry of Education. (1988). Curriculum Management. Toronto: Ontario Government Publications.

Patel, V.L., & Groen, G.J. (1986). Knowledge based solution strategies in medical reasoning. Cognitive Science, 10, 91-116.

Penner, B.C. & Voss, J.F. (1983). Problem solving skills in the social sciences: Methodological considerations. University of Pittsburgh: Learning Research and Development Center.

Perkins, D.N. & Salomon, G. (1989). Are cognitive skills context-bound? Educational Researcher, 18(1), 16-25.

Rutherford, W.L. (1985). School principals as effective leaders. Phi Delta Kappan, 67, 31-34.

Stiggins, R. J. (1988). The nature and quality of teacher-developed classroom assessments. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Stiggins, R. J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22 (4),271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5 (2), 5-17.

Terwilliger, J. S. (1987). Classroom evaluation practices of secondary teachers in England and Minnesota. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D.C.

Traub, R., Nagy, P., MacRury, K., & Klaiman, R. (1988). Teacher assessment practices in a senior high school mathematics course. Final Report, Transfer Grant 52-1028. Toronto: The Ontario Institute for Studies in Education.

Voss, J.F., Greene, T.R., Post, T.A., & Penner, B.C. (1983). Problem solving skill in the social sciences. In G.H. Bower (Ed.). The psychology of learning and motivation: Advances in research theory. New York: Academic Press.

Wahlstrom, M.W. & Danley, R.R. (1976). Assessment of student achievement. Toronto: Ministry of Education.

Table 1
Matrix of Activities

Principal	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Student															
academic info	1			4	5			8			11		13		
personal info	1			4	5						11	12	13		
academic assistance							7	8		10					
personal assistance				4	5					10	11		13		15
direct contact	1	2									11	12	13		15
learning style							7					12		14	
Teacher															
competence info	1							8							
attitude info		2			5			8				12			15
study-like info		2						8		10	11		13		15
academic assistance				4				8		10					
personal assistance				4									13		
flexibility info				4						10					
other teacher info										10					
Department															
study-like data			3	4					9	10	11		13		15
promote conformity	1	2	3	4					9			12		14	
don't promote						6									
use one teacher							7					12			
flexibility info										10					
tutorial set-up	1														
success info						6									
dismiss report												12			
School															
transfer info	1				5										
department info															15
contain problem				4	5					10					
Parent															
meet with principal	1	2	3		5								13	14	15
meet with teacher				4		6	7		9						
discourage move			3	4					9		11				
assess if vocal				4											
ongoing talk										10					
dismiss complaint														14	
District															
admin view - study		2													
avoid system exams				4											
ignore trustee						6									
system committee							7								
system policy				4											
Other															
general strategy								8							

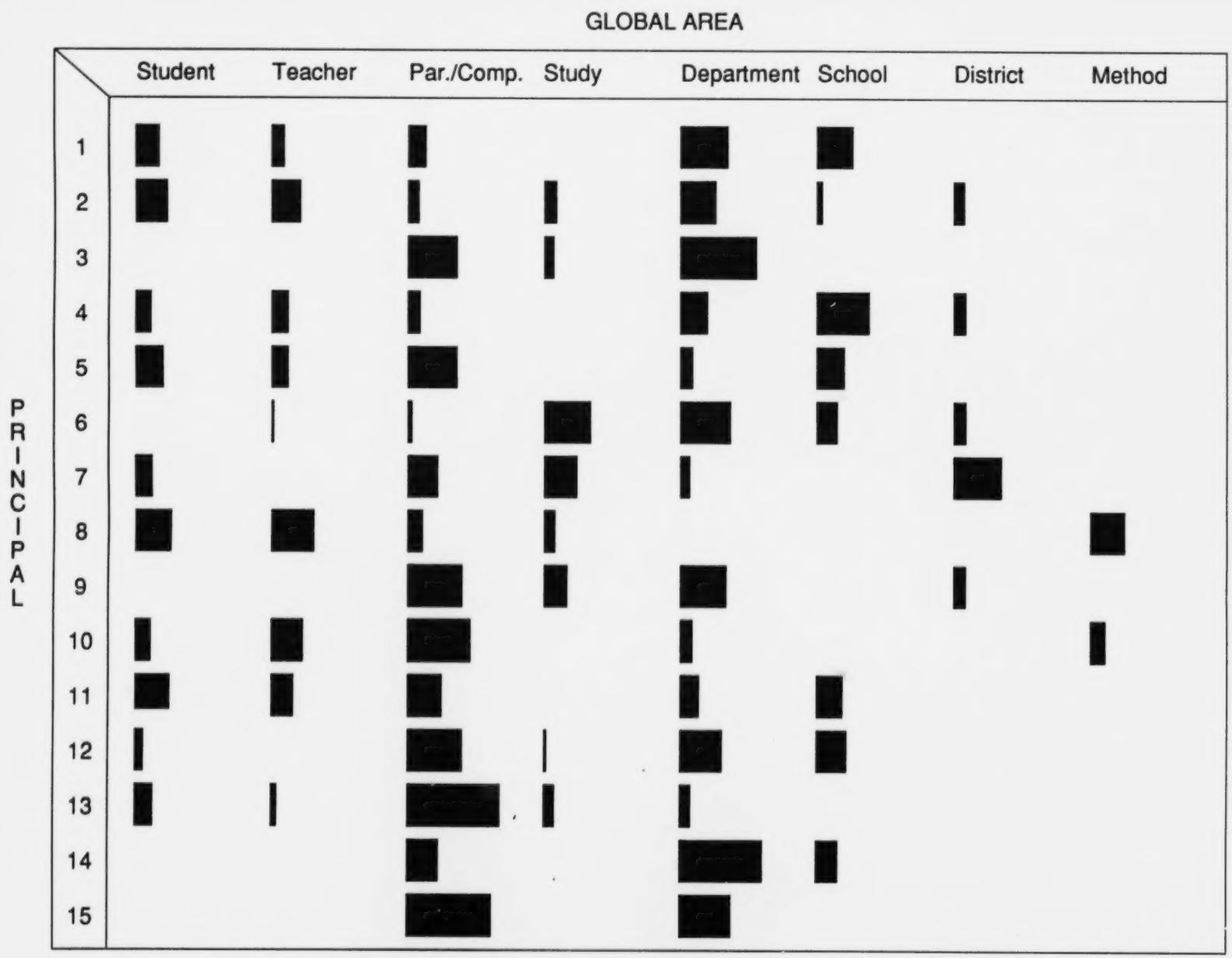

Figure 1: PERCENTAGE STATEMENTS BY GLOBAL AREA

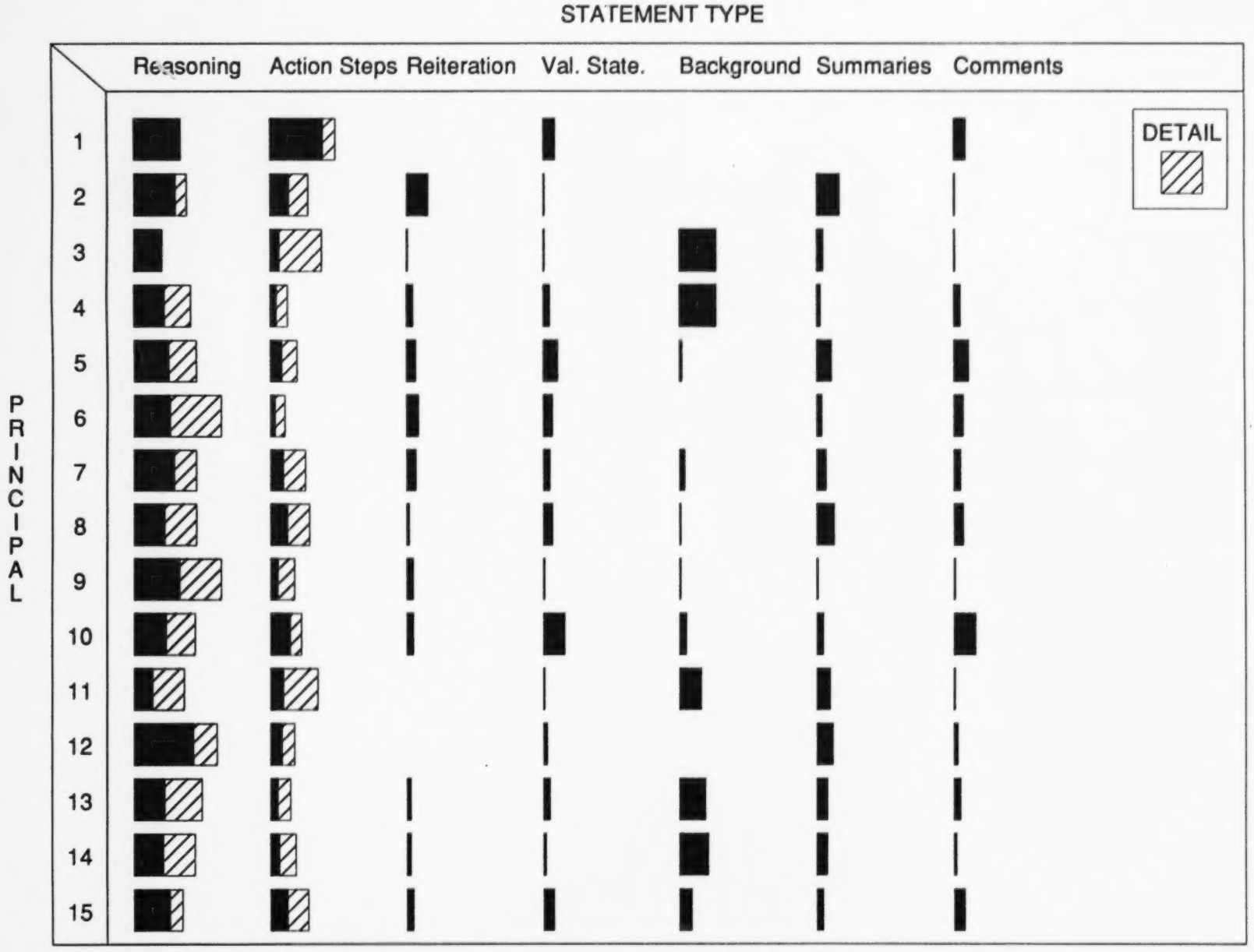

Figure 2: PERCENTAGE STATEMENTS BY TYPE

APPENDIX A

STUDENT EVALUATION CASE STUDY

BACKGROUND

Smithson Collegiate Institute serves a predominantly affluent, professional suburban community; however, there is one feeder school in a lower middle class neighbourhood with subsidized housing. There are just over 1400 students, substantial numbers of whom go on to university after graduation. Mr. Lesley is a recently appointed principal to Smithson Collegiate, just beginning his second year.

A district research study on student evaluation in calculus classes has just been completed, and a summary sent to each secondary school, along with a memo from the assistant superintendent of program requesting that schools review the results and their implications. An abbreviated summary of the 80-page report is appended.

THE ISSUE

When the results of the study were presented at a board meeting, there was considerable discussion. One influential and experienced trustee stated that he expected that the situation was even more serious among less experienced teachers and in more difficult to evaluate subject areas.

Mr. Lesley sent this information to the math department head, Mr. Addison, who has held that position for 11 years, the past six at Smithson Collegiate. There are four calculus classes in the school, with four different teachers. The calculus course coordinator is a senior teacher with an excellent reputation among his peers for the quality of his teaching. The other three have several years of experience and are regarded as being competent. Mr. Addison told Mr. Lesley that he was not aware of any problems such as those mentioned in the report. Mr. Lesley asked him to raise the issue at the next department meeting and to keep him informed. The head agreed.

All department meeting minutes are routinely sent to Mr. Lesley, and the next set for the math department meeting included an item on the calculus study. Mr. Lesley noted that after some discussion, the teachers involved felt that the problems reported were not significant at S.C.I., since there was a common calculus exam and a policy of no exemptions. The department head appended a note saying that, although there were differences of opinion re teaching emphases, he felt strongly that the calculus classes were well taught by experienced staff members, each of whom had individual strengths, and that any attempts to impose greater uniformity would be counter-productive.

A week later, an angry parent demanded to see Mr. Lesley about the calculus class his daughter was enrolled in. The teacher, he said, had a negative attitude towards all the students who weren't especially talented in mathematics. His daughter was an "A" student who worked hard and had always achieved well, but was finding calculus difficult. The parent accused her teacher, Mrs. Peerless, of not having enough patience with those students who needed more time for practice and review of the concepts. The parent was indignant and demanded that his daughter be switched to another calculus class

taught by Mrs. Lovejoy, who had a reputation for being enthusiastic about the subject and whose students always seemed to do well.

You are Mr. Lesley. How will you handle the situation?
Please think out loud as you consider this situation and your reactions to it.

REPORT SUMMARY

The purpose of the report was to explore the meaning of grades assigned by different teachers of the same calculus course. The participants were 15 teachers, one from each high school in the board. All participants were highly qualified and experienced teachers of the course. The data collected were as follows:

(1) a log giving the classroom activities each day and the time devoted to each;

(2) a list of the homework and seatwork assigned each day;

(3) a list of the criteria used to arrive at student grades for the course; and

(4) grades assigned by the teachers to a common set of 20 final exams obtained from a class not involved in the study.

Great differences in practice were found.

Among the teachers, the amount of homework and seatwork assigned varied greatly. When all activities listed in the logs were classified into one of six categories, variations in time spent in each activity were as follows:

administration, from 0 to 5%;
direct teaching, from 17-52%;
student practice, from 8-47%;
homework, from 11-43%;
review, from 4-14%; and
assessment from 8-16%.

Teachers who placed a relatively high emphasis on practice activities tended to place a relatively low emphasis on review and assessment activities. Moreover, those with more class time available expended a smaller percentage of time on direct instruction, and allocated a greater percentage to homework and practice.

From the reports of work assigned, the results showed that teachers varied widely in the number of questions assigned as homework--from under 500 to more than 1600. Substantial differences were also found among teachers in content emphasis. For example, the number of questions on basic skills ranged from 40% to 75% of the total number of assigned questions; the number on proofs from 0% to 18%. Similar variations were found in the other content groups. When content of what was taught and what was tested were compared, there were irregular patterns in topic emphases. The emphasis on basic skills was less on

the tests than the assignments, whereas the emphasis on other content groups was greater on the tests than on the assignments.

The analyses of the grading practices revealed large differences in the type of evidence collected by teachers for their grades. Examinations and term tests were the two main determinants of student grades. For the four classes (of 15) following an exemption policy, the majority of students took their only exam on material learned in the first half of the semester. For six other classes, the only exam was a final exam based on the entire semester's work. In the remaining classes, both a mid-term and final exam were required. Across teachers, the final examination mark was weighted from 15% to 40% of the student's final grade and, when a mid-course exam was administered, the resulting mark was weighted 9% to 30% of the final grade. Work in the first and second halves of the semester were weighted equally by five of the teachers, and the second half was weighted more heavily (70%) by the other ten.

The study of teacher marking of the common set of 20 exams revealed widely different expectations of student performance for marks awarded, as seen in substantially different marking schemes for the same questions and in the existence of 'easy' and 'hard' markers. Although there was substantial agreement among the teachers as to the relative importance of the examination questions and the relative quality of the student papers, there was substantial disagreement as to the absolute quality of the student papers. To illustrate, three teachers assigned no paper a mark in the honours range, and one teacher assigned failing marks to seven papers; on the other hand, seven teachers assigned no paper a failing mark, and one teacher assigned marks of 80 or more to 10 of the 20 papers.

APPENDIX B

VALUE STATEMENTS

Principal 1:

Department--promote conformity

If not, they should be (common tests, exam), in other words one teacher sets the first test, another one sets the second one, and so on.

Principal 2:

Department--promote conformity

It (the school) should (have an overall evaluation policy).

Department--study-like data

That isn't good enough (accept department's assessment).

Department--promote conformity

I don't want one class doing 8 percent and another class 40%.

Principal 3:

Department--promote conformity

My particular philosophy is that parents and students have a right to feel that in one particular course where there are three or four or five different instructors that there is some continuity and some consistency of practice and evaluation.

Department--promote conformity

While we are not trying to make robots of the teachers, we do feel that similar courses with similar evaluation practices should more or less resemble each other.

Department--promote conformity

And while there will be some small difference in judgement, there is professional judgement, but that judgement should not be completely out of wack or based on entirely different tests that were given or different methods of evaluation.

Principal 4:

School--transfer info

I think, first of all, you undermine confidence in your teachers, period, if you're willing to move them around like players on a chess board.

Teacher--info

I think that's a principal's responsibility to see that the teachers' doing what they're supposed to be doing.

Other

I'm not a believer in, first of all, undermining the teaching staff's strengths, period.

Student--personal assistance

I am not a fan of personality conflicts. I don't allow it... It's not something I encourage.

District--system policy

We feel different about final exams than they (other schools in board) do and I think you have to allow a certain amount of flexibility, but there has to be some check.

District--avoid system exams

You don't need examinations, maybe (in a school where focus is on business or tech) and you certainly don't need standardized testing in calculus for the number of kids you are dealing with.

District--avoid system exams

And I think you ought to keep that independence in a school.

District--avoid system exams
I think you ought to be able to be able to address the needs of your kids.
Other
The grade 9 and 10 teachers should be dealing with the elementary feeder schools.
Department--promote conformity
And I think this common set of exams, again it (discrepancy) shouldn't happen within a school.
Department--promote conformity
There should be a marking scheme that's common.
Department--promote conformity
There should be a common exam for the four sections in the school, if there are four teachers.

Principal 5:
School--contain problem
And what you always want is to have the first avenue of appeal within the school.

Principal 6:
Department--promote conformity
My feeling is that I would expect to see a relatively high degree of commonality in such things as courses of study, expectations of students, evaluation techniques, textbook being used, what is the purpose of evaluation, what methods are used, how much sharing of evaluation techniques, right through the whole year.

Principal 7:
Other
I don't think you're going to get much support or much trust from teachers of any subject level working together to bring about betterment for the system if you're going to compare one school to another, or for that matter, even teachers within that school.

Principal 8:
No value statements

Principal 9:
Department--don't promote
I don't think the solution is to demand that the daughter be switched to another calculus teacher.
District--system policy
The department heads across the county should have some broad general standards...
Department--study-like data
and the department head in this particular school, when slicing that calculus pie into four pieces, should have some internal way to monitor the stresses in the courses--do we stress seatwork, do we stress the daily testing, do we have any assessment as opposed to evaluation, do we have make-up opportunities?...master learning?...chance to do (test) again?
Department--promote conformity
There should be some understanding between these people so that we eliminate many of these differences...

Principal 10:
Other
I think you try and treat teachers...treat them with respect.

Principal 11:
No value statements.

Principal 12:
Student
Whatever decision is made would have to be made in the best interest of the student, regardless of how the parent feels, and regardless of where the parent is coming from, and regardless really of what the teachers feel.

Principal 13:
Teacher--attitude info
Sometimes that (teacher not being sensitive) happens knowingly and that probably bothers me more than anything else.

Principal 14:
Department--promote conformity
I think in a common public educational system we need to be very very much aware of this and do whatever we possibly can to reduce that variance that could occur.
Department
You'll never, I don't think, get exact uniformity 'cause there are different teaching styles there...

Principal 15:
Other
I believe that no matter where we are we can continue to improve and make things more effective.
Department
I don't feel that everybody has to teach lock-step, that we're all doing all the same thing all of the time, because I think that certainly stymies individuality.
Department--promote conformity
I also believe that there's got to be a core, that we are doing basically a similar thing, especially with respect to evaluation...
Promote conformity
I think it's only fair that there has to be a degree of continuity not only within a school but even within a system.
Department--don't promote
I believe that we can't run away from a timetable from one person to the next.
Student
I believe after thorough investigation if we have a personality conflict or a learning situation that is being stymied, then perhaps a change may be possible.